Design Alternatives to AI Proctoring Software

by

Aditi Kumar

Integrated MA, Economics (2013)
Indian Institute of Technology, Madras

Submitted to the Integrated Design and Management Program
in Partial Fulfillment of the Requirements for the Degree of

Master of Science in Engineering and Management

at the

Massachusetts Institute of Technology

June 2021

Signature of Author .....
Integrated Design and Management
May 14, 2021

Certified by .....
Tony Hu
Academic Director, Integrated Design and Management, MIT
Thesis Supervisor

Accepted by .....
Matthew S. Kressy
Executive Director, Integrated Design & Management Program

This page is intentionally left blank

Design Alternatives to AI Proctoring Software

by

Aditi Kumar

Submitted to the Integrated Design and Management Program

On May 14, 2021 in Partial Fulfillment of the Requirements for the degree of

Master of Science in Engineering and Management

Abstract

With the increasing popularity of and avenues for online learning, a considered approach to academic integrity has emerged as a priority. Artificial intelligence (AI) enabled proctoring software has been touted as a solution. However, it relies on imperfect and biased technologies: face detection, facial recognition, and gaze detection. In the real world, or in the "wild", as it is referred to in computer vision, these technologies can underperform for darker skin tones in poor light, poor internet connectivity, amongst other things, and surface several false positive "flags" for review per student, thereby increasing the operational costs at scale. These have also been shown to create an atmosphere of stress for students. Through primary research that included analysis of internal data at an education technology company and stakeholder interviews, I identified key needs relating to academic integrity. I then recommended alternatives to AI proctoring that include assessment design and administration, honour code, and in-house authentication. In addition to recommending the alternatives, a phased implementation plan was also created.

Thesis supervisor: Tony Hu

Title: Academic Director, Integrated Design and Management

This page is intentionally left blank

Table of Contents

Introduction

In the wake of the covid-19 pandemic that struck human society towards the end of the 2019, we fully realized and experienced how the internet had gone from becoming a luxury to a necessary infrastructure in modern society. In the prevailing atmosphere of uncertainty about how the disease spread, and therefore how one could safeguard oneself, almost overnight, all the establishments, shops, offices, and schools as physical spaces had to be cordoned off. Several articles called it an acceleration of future trends[1]– virtual, remote work, online education, and so forth. The following chart from the Mckinsey report cited here demonstrates this acceleration based on expert interviews:

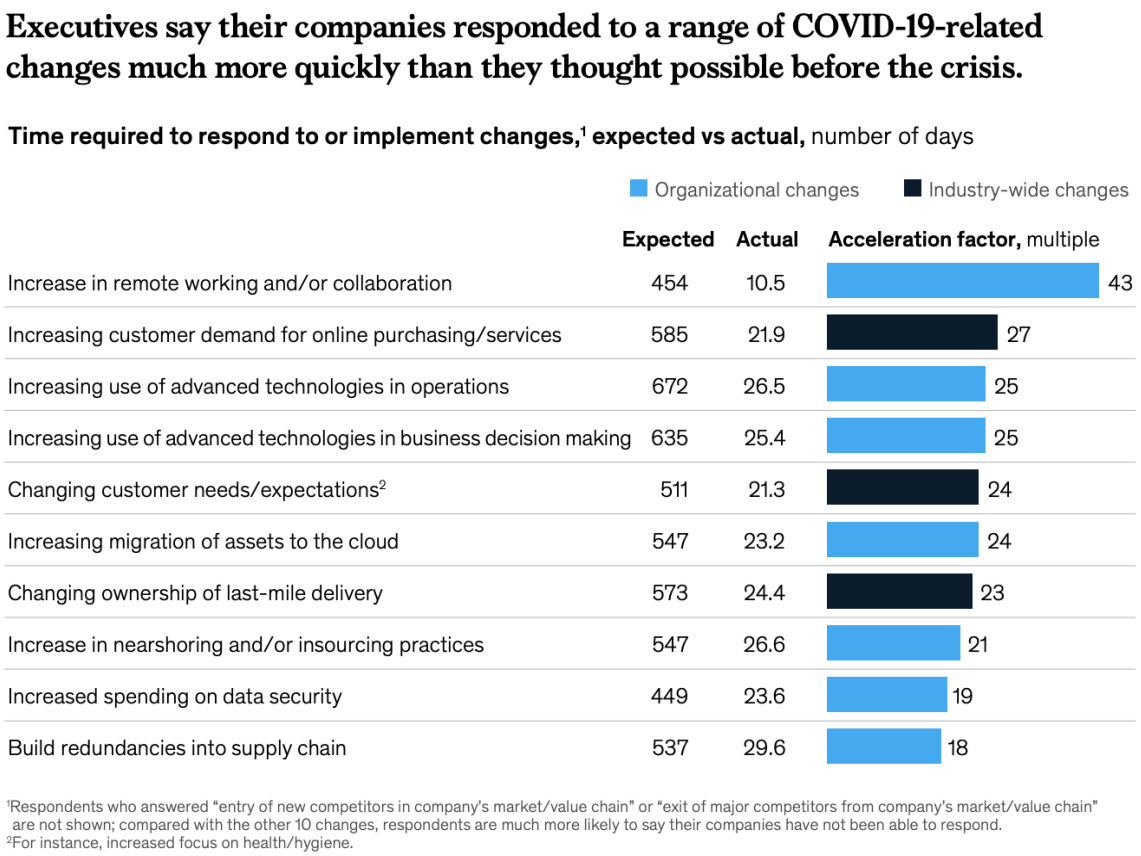

Fig.1: Mckinsey & Co. Exhibit showing acceleration of technological trends in the wake of covid-19 pandemic

[1] "*How COVID-19 has pushed companies over the technology tipping point—and transformed business forever*", Survey, McKinsey & Co., October 2020

Long before covid-19, the education sector had already seen an increase in the use of the internet as a distribution channel for lifelong and adult learners. Lifelong learning is a construct that is a simple yet profound idea– that learning starts the moment a person is born, continues throughout their life. However, education is structured in the modern world as arising from and granted by formal institutes. The idea that we as a society need to create opportunities for learning that are recognized universally in support of the idea of lifelong learning is therefore a powerful one. In this context, it is useful to surmise the basic precepts of lifelong learning, as identified by Knapper (1985):

- available throughout an individual's lifetime;
- would respond to each person's needs to cope with the demands of contemporary society;
- would involve learners in guiding and directing their own learning;
- and would encourage learning from a variety of sources, both formal (for example schools) and non-formal (for example in the workplace or from colleagues).

The proliferation of personal computing devices and the internet has indeed given a fillip to the mission of lifelong learning.

In this dissertation, I will be considering online learning as a subset of online distance learning, similar to Conrad et al (2018). I will refer to covid-19-related online classes as remote learning. Online learning can cater to various needs, viz. academic credentials, supplemental learning, professional development, skill-building, and so forth. It can therefore lead to various outcomes for the end-user such as certificates, micro-degrees, recognized degrees or credits. Online courses can

require either synchronous or asynchronous learning. Synchronous online learning occurs when the instruction and the attendance co-occur in an internet-based medium, e.g. students attending a live online lecture in real-time. In asynchronous online learning, the instruction and attendance are decoupled, e.g. students viewing pre-recorded content. Similarly, courses can be hybrid too, where instruction is spread across online and in-person, thereby mixing traditional classrooms with online learning. Another dimension of online learning is whether it is individual or cohort-based. Individual study programs and courses allow the learner to enroll at any point during a year, and beyond that, it may be self-paced or of a fixed, stipulated duration. Cohort-based courses allow enrollment in batches, as there may be a community or collaborative learning element such as a discussion board or group projects. These are also almost always of a fixed, stipulated duration.

For this dissertation, I worked with Outlier.org, an education startup providing cohort-based, asynchronous, college-level courses for credit in partnership with University of Pittsburgh, recognized by the accreditation rules prescribed by the US Department of Education. For distance education, the criteria of evaluation for accreditation need not be different, however, the *Accreditation Handbook 2020*[2] published by the Accreditation Group in the Office of Postsecondary Education of the U.S. Department of Education, states that:

"The [accrediting] agency meets this requirement if the agency demonstrates that it--

...(g) Requires institutions to have processes in place through which the institution establishes that a student who registers in any course offered via distance education or correspondence is the same student who academically engages in the course or program; and

[2] Accreditation Handbook, U.S. Department of Education Office of Postsecondary Education Accreditation Group

(h) Makes clear in writing that institutions must use processes that protect student privacy and notify students of any projected additional student charges associated with the verification of student identity at the time of registration or enrollment."

This poses an important requirement to the design of a learning product like Outlier. Additionally, online courses have found it nearly impossible to do away with high-stakes testing and assessments. In high-stakes testing, a major portion of the grade is assessed in a single examination, such as a final examination. In a traditional classroom, high-stakes examinations are proctored in person by the teaching team. In online learning, exam invigilation is a thorny issue that has been sought to be solved via remote proctoring. Invigilation on exams has been upheld as a tradition across formal institutes because academic dishonesty, whereby a learner accesses forbidden materials or help to answer test questions, is taken as a foregone conclusion. As Holden et al (2020) state in their review of the literature around cheating and online learning, "Although there is disagreement within existing research as to whether or not more cheating actually occurs in online exams compared to in-person exams, there is agreement that regardless of delivery method, cheating has always occurred".

Design Brief

After discussing with the internal teams, the following brief was finalized for this project.

It would help us if you'd research alternatives to remote online proctoring and come up with a recommendation (or recommendations) for potential future implementation. Essentially, we'd look for you to help us understand:

- *How could we design an experience that discourages or prevents students from cheating without using remote proctoring?*
- *What trade-offs might these experiences require?*
- *How could we test whether these solutions actually do discourage or prevent students from cheating on exams?*

We'd look for you to provide us with a body of research and design/product recommendations. We may then review your recommendations with our stakeholders to decide whether to move forward.

In the following section, I will elaborate on the core constructs that are under consideration in this dissertation.

Constructs

Academic Integrity and Dishonesty

Holden et al (2020) provide a comprehensive, usable definition of academic dishonesty:

"Academic dishonesty or "cheating" includes behaviors such as the use of unauthorized materials, facilitation (helping others to engage in cheating), falsification (misrepresentation of self), and plagiarism providing an unearned advantage over other students. Broadly, the behaviors are not consistent with an established University's Standards of Conduct, which communicates expected standards of behavior. "E-dishonesty" more specifically refers to behaviors that depart from academic integrity in the online environment. Even more specifically, concerns in relation to online exams typically include 'electronic warfare' (tampering with the laptop or test management system), impersonation, test item leakage, and use unauthorized resources such as searching the

internet, communicating with others over a messaging system, purchasing answers from others, accessing local/external storage on their computer, or accessing a book or notes directly."

With this definition, we see that academic dishonesty consists of 3 distinct behaviours, to wit:

1. Use of unauthorized materials, e.g. using a textbook in a closed book examination
2. Plagiarism
3. Misrepresentation of self/ impersonation, e.g. getting a friend or hired help to take a test in your place, which I will be referring to as "authentication" in this dissertation
4. Leaking questions to other students/ on the internet

Ultimately, academic dishonesty endangers the credibility of the grade awarded to the student, and over time, this might lead to reputational as well as material (loss of accreditation) damages to the institution.

Academic integrity can be understood as the prevention of academic dishonesty as well as promotion of the culture of honesty and responsibility, and increasingly, "an approach that is educative, preventative, and positive in promoting student success"[3]. There may be a number of reasons why learners engage in academic dishonesty– systemic (university culture, university policies), individual factors (stress, motivation, fear, etc.), and situational (a combination of systemic and individual factors aligning with one another).

[3] From Berkeley Center for Teaching and Learning

The question of cheating and authentication becomes even more salient in high-stakes examinations. Commonly, examiners have used remote proctoring, passwords, and randomized question sets to deter and/or detect cheating.

Proctoring

In the world of distance education and remote examinations, the question of invigilation is a very salient one. As discussed elsewhere in this dissertation, proctoring is done with the assumption that academic dishonesty would occur, and therefore render the integrity of the grade or credit unreliable. Today, there are broadly 3 types of proctoring:

1. In-person: a member of the teaching team or institute/ hired proctor is present in the examination hall where the test-takers are also seated. Here, the environment is controlled, and the proctor looks for suspicious behaviour.
2. Remote human proctoring: a member of the teaching team or institute/ hired proctor logs into the same web conferencing room as the test-taker. The test-taker may be required to "show" the room via webcam for a room scan. The proctor monitors the video and audio feed of the test-taker for any suspicious activity. The task of finding a proctor (for each individual student) may be outsourced to a proctoring agency, or the teaching team may choose to proctor remotely.
3. Remote AI proctoring: usually a client (such as a higher education institute or university) buys the AI proctoring services from a vendor. The test-takers are usually required to download a browser plugin which controls various functions of the camera, microphone, keyboard, and browser, depending on the settings chosen by the client. The AI proctoring

softwares claims to have proprietary algorithms running under the hood, which identify various flags in the video/ audio feed for someone in the client team to review and take action on. These algorithms are usually face detection (similar to the one found on an Android or iPhone camera, a rectangle around the human face in the frame), facial recognition (a matching of the face in the video frame to a catalogue of images or a single image like a student ID uploaded by the student), or gaze detection (observing the movement of the eyes of the user with respect to a fixed frame such as a laptop screen in front of them). The issue with these technologies could be many:- a) the algorithm could be trained on the most commonly found faces, thereby not accounting for diversity, also known as edge-cases in machine-learning-speak, b) the algorithm may perform poorly at detecting darker skin tones in poor light, as the edge detection may work poorly, c) facial recognition may fail if the student ID has an unclear or old photograph, or if the test-taker is dark-skinned and in poor lighting, or if there is a change in facial hair, weight, and other physical attributes, or a combination of these factors, d) poor internet connectivity could make the video feed less reliable and therefore turn up many flags or none (Selinger and Leong, 2021).

In addition to these issues, it also treads the thin line of surveillance and monitoring, leading to questions about (perceived or otherwise) breach of privacy.

One of the biggest questions that researchers, learning design professionals, and teachers have had to consider is the difference in student performance, time taken, and possibility of academic dishonesty in the context of unproctored versus proctored examinations online. The literature speaking to this question, and to the efficacy of proctoring has seen a big surge in the past decade,

in line with the growth in enrollment in distance education. There is no conclusive evidence as to whether or not cheating is higher in unproctored examinations as compared to proctored examinations (Holden et al, 2020).

Assessment

Erwin (1991) defines assessment as the following, “Assessment is defined as the systematic basis for making inferences about the learning and development of students. More specifically, assessment is the process of defining, selecting, designing, collecting, analyzing, interpreting, and using information to increase students' learning and development”

Evaluation, as distinct from assessment, usually refers to the evaluation of the efficacy of the learning design and the course overall. Taken together, both look to ultimately serve the goal of improving learning outcomes.

Assessment, in terms of how it is administered, is of two types:

- Summative: which occurs at the end of a period of learning, e.g. mid-term or final examinations
- Formative: which occurs continuously through the period of learning, e.g. in-class reflection activities, learning journals

Sullivan (2016) called for an integrated approach to tackling academic dishonesty in online learning which relied heavily on a scalable, inventive assessment strategy that diversified the type of questions and administration of questions with the help of technology.

Similarly, technology can be in service of the chosen assessment philosophy and design. Ideally, learners must know what they are expected to be tested on, and the rubrics for formative

assessment must also be made clear upfront. Designing assessments backwards from the learning outcomes is the cornerstone of "backward design" in instruction design (McTighe and Wiggins, 1998). The following is a framework for assessment from the resources accompanying the book, "Understanding by Design" by McTighe and Wiggins:

Classroom Assessment Planning: Key Questions

Content Standards	Purpose(s) for Assessment	Audience(s) for Assessment
What do we want students to know, understand, and be able to do?	*Why are we assessing and how will the assessment information be used?*	*For whom are the assessment results intended ?*
■ _____ _____ _____ ■ _____ _____ _____ ■ _____ _____ _____ ■ _____ _____ _____	❑ diagnose student strengths and needs ❑ provide feedback on student learning ❑ provide a basis for instructional placement ❑ inform and guide instruction ❑ communicate learning expectations ❑ motivate; focus student attention and effort ❑ provide practice applying knowledge and skills ❑ provide a basis for evaluation __ grading __ promotion/graduation __ program selection/admission ❑ provide accountability data ❑ gauge program effectiveness	❑ teacher/instructor ❑ students ❑ parents ❑ grade-level/department team ❑ other faculty ❑ school administrators ❑ curriculum supervisors ❑ business community/employers ❑ college admissions officers ❑ higher education ❑ general public ❑ other: _____

Fig. 2: A "backward design" template for planning classroom assessment

Framework of Assessment Approaches and Methods

How might we assess student learning in the classroom?

SELECTED RESPONSE ITEMS	PERFORMANCE-BASED ASSESSMENTS			
	CONSTRUCTED RESPONSES	PRODUCTS	PERFORMANCES	PROCESS-FOCUSED
❑ multiple-choice ❑ true-false ❑ matching	❑ fill in the blank • word(s) • phrase(s) ❑ short answer • sentence(s) • paragraphs ❑ label a diagram ❑ "show your work" ❑ representation(s) • web • concept map • flow chart • graph/table • matrix • illustration	❑ essay ❑ research paper ❑ log/journal ❑ lab report ❑ story/play ❑ poem ❑ portfolio ❑ art exhibit ❑ science project ❑ model ❑ video/audiotape ❑ spreadsheet	❑ oral presentation ❑ dance/movement ❑ science lab demonstration ❑ athletic skills performance ❑ dramatic reading ❑ enactment ❑ debate ❑ musical recital ❑ keyboarding	❑ oral questioning ❑ observation ("kid watching") ❑ interview ❑ conference ❑ process description ❑ "think aloud" ❑ learning log

Fig. 3: A "backward design" framework for organizing classroom assessment types

Case Facts and Needs

Outlier is an education technology company, launched in 2019, that seeks to help students offset education costs and debt by aiming to provide high-quality, engaging content at a price lower than a traditional university. Outlier has partnered with University of Pittsburgh to grant transferable credits to learners upon successful completion with satisfactory performance on examinations. This implies that Outlier is subject to the high standards of the national accreditation body in the United States, under the US Department of Education. This is distinct from other higher education learning services such as Coursera, which provides certificates, but not recognized credits. For this reason, Outlier as an organization is committed to ensuring high standards of academic integrity– the value and integrity of the credits granted to the student must remain intact.

When the covid-19 pandemic hit in 2020, Outlier found itself in a position of high demand, as all universities went online, and the playing field, as it were, was levelled. Outlier provides the option of taking a 7-week or 14-week introductory, cohort-based, college-level course, for subjects like Calculus, Philosophy, Psychology, Microeconomics, and Astronomy. Completion of a course grants 3 transferable credits. At roughly $130 per credit, Outlier is nearly three times cheaper than comparable credits at a standard university. The course is administered through a Learning Management System (LMS) custom-built internally at Outlier.

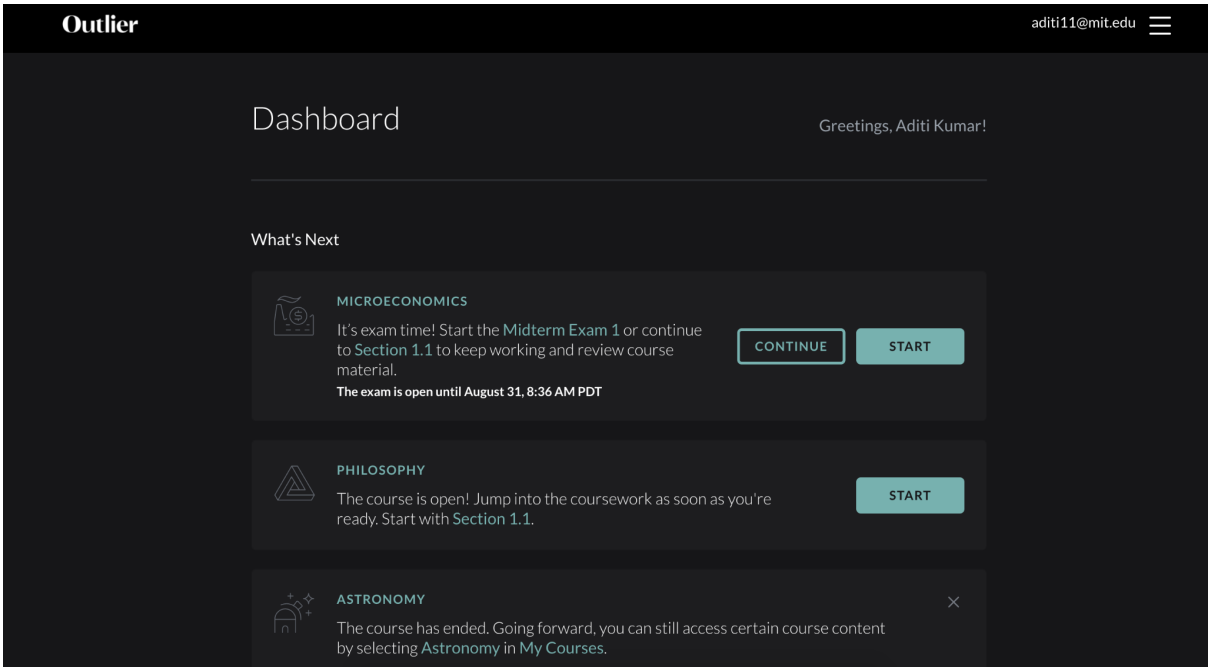

Fig 4:. User-facing LMS Dashboard at Outlier

Each course consists of multiple chapters, and sometimes, sub-chapters as well. Each (sub-)chapter consists of roughly 4 to 5 video lectures, of 8 to 10 minutes each, an e-textbook, short quizzes, and practice problems. The course also consists of 0 to 2 midterm examinations and 1 final examination. The quizzes, midterm examination, and the final examination count towards the final letter grade. The proctored examinations count for anywhere between 30% to 50% of the final

grade. Most questions are multiple choice, fill-in-the-blanks, or true/false. Introduction to Philosophy consists of essay questions as well.

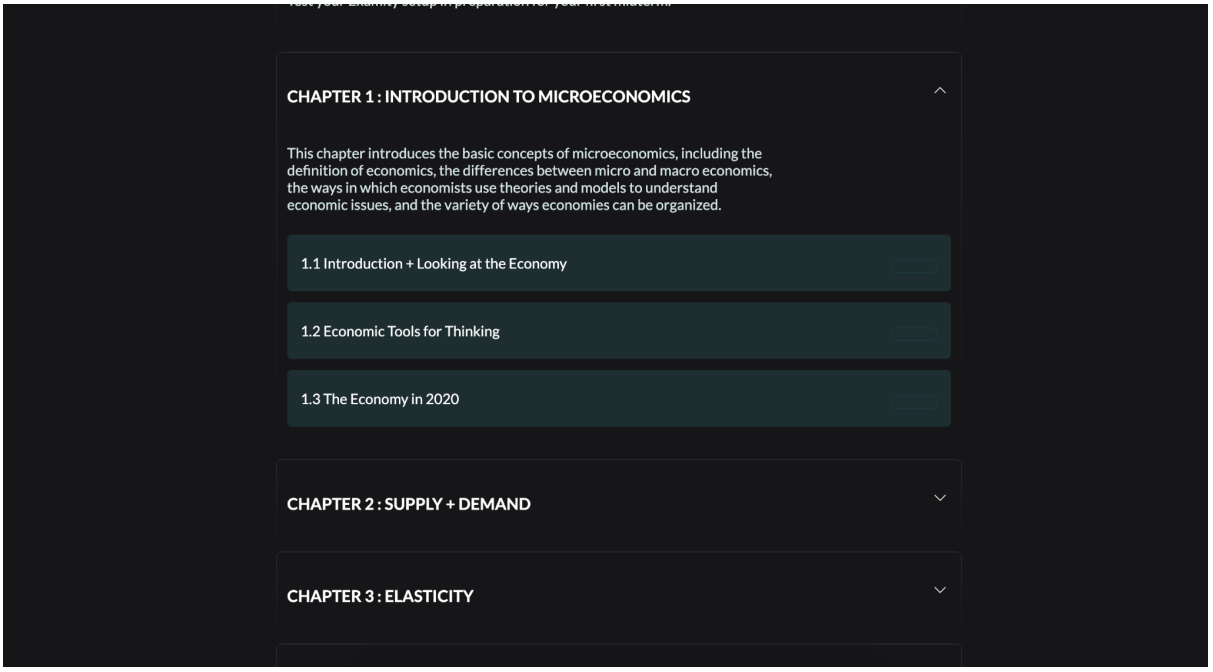

Fig. 5: The course site for Microeconomics

;

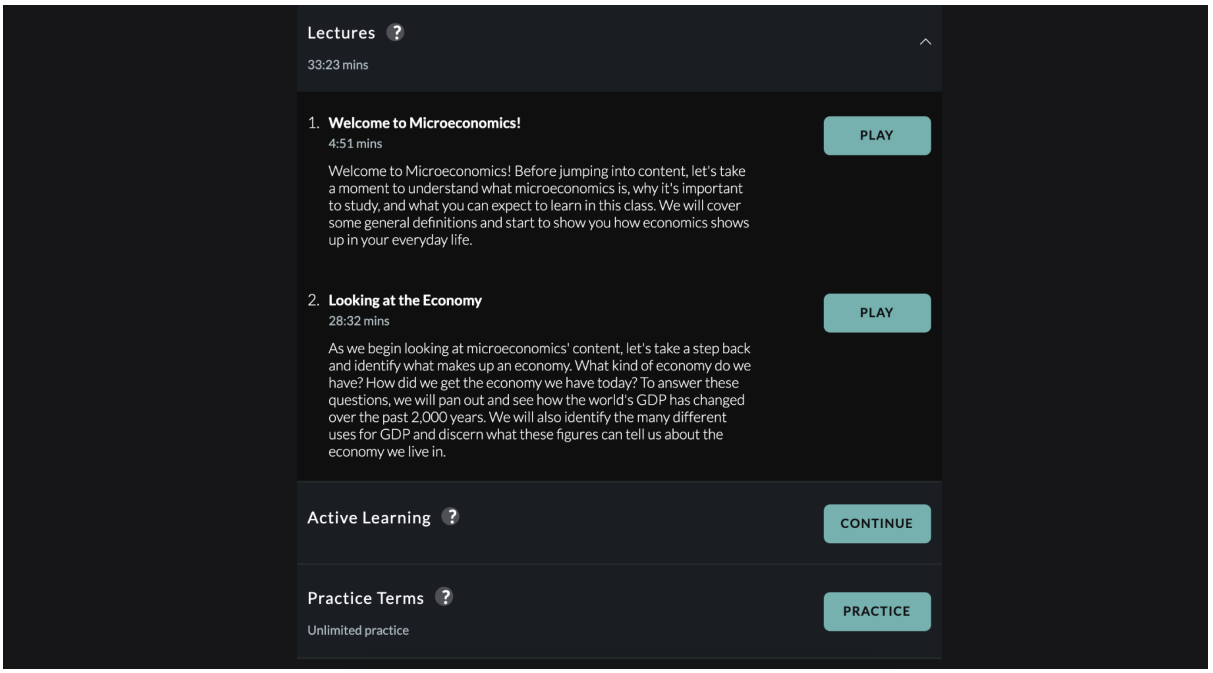

Fig. 6: The items within a Chapter 1 of Microeconomics

The Outlier team consists of the Content, Product, and Student Success teams, amongst others, and these three teams have a direct interaction or influence on the product design, development, and stakeholder management. A user can purchase the course from their website, and set up the login for the LMS dashboard.

Content-related queries that students have are routed to either the tutoring or the content-team (which includes instructors), depending on the nature of the question. Students also participate on a discussion board via Slack. The student success team is actively present on Slack, and handles all the administrative queries that the students have. For all complaints lodged by students (either via email or via Slack), the student success team creates tickets on ZenDesk. As the questions are resolved, the tickets are also resolved on ZenDesk. Tickets are also tagged; some examples of tags are “exams_general”, “final_grade”, “academic_integrity”, etc.

The proctoring is delivered via a third-party AI-based proctoring software. Amongst the AI proctoring softwares commonly used in online learning, Proctorio claims that they only use facial detection and eye-gaze detection software, and not facial recognition. It is important to note that facial and eye gaze detection softwares can easily fail in low-light situations for black and brown people especially. AI proctoring software platforms typically come with multiple product features, which the test-administering institute can choose from. These may include (Hussein et al, 2020):

- Audio recording
- Keystroke monitoring
- Browser lockdown

- Video recording pause/ restart options
- Disable cut-copy-paste function
- Hide taskbar
- Hide desktop
- Prevent right-click
- Prevent webpage navigation (forward/ backward)
- Keystroke analytics for authentication
- Face detection
- Face recognition (against a student ID, say)

Outlier has utilized only some of the above features, including video and audio recording. There are some differences between the various companies, but essentially, as proctoring vendors, they provide flags for the client team (Outlier's student success team in this case) to examine, and leave the decision-making of whether or not an infraction was committed to the client team. The flags are created automatically via the proprietary algorithms, based on the video stream of the test-taker. The client team typically reviews the video recording associated with the flag to determine if they wish to follow up with the student or not. This may violate student-centricity, as there may be clients who make decisions without bringing the student into the fold or informing them. Thus these softwares could potentially be misused, as the bias of human decision-makers is not solved for. The student video and audio recording is deleted after a certain period of time.

The following image is an example of the user journey for a student with respect to the final examination experience. Again, there may be variations in the login and/or setup process from one

proctoring platform to another, but the point to note is that the examination always takes place on a separate learning platform (e.g. LMS in use at the university).

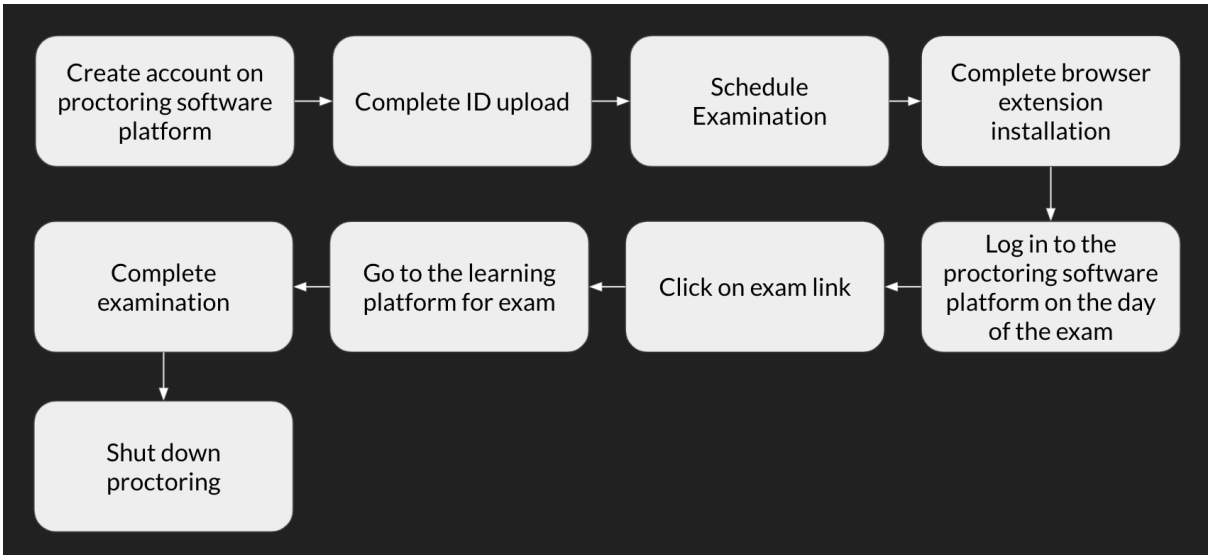

Fig. 7: User journey for high-stakes examination

The student success coordinator in charge ends up spending 1-2 full days on examining the flags thrown up by the proctoring software (which could be anywhere between 10 and 60 per student). Of these flags, typically about 5% to 8% require further investigation and creation of tickets to be resolved. An example of the gradebook from a proctoring software is below, with the multiple flags showing up as yellow and red spikes[4]:

[4] Accessed from Brigham Young University Office of IT Support website on May 8, 2021 (https://support.byu.edu/it?id=kb_article&sys_id=19f727b2dbaf48141f061cb51b96198a)

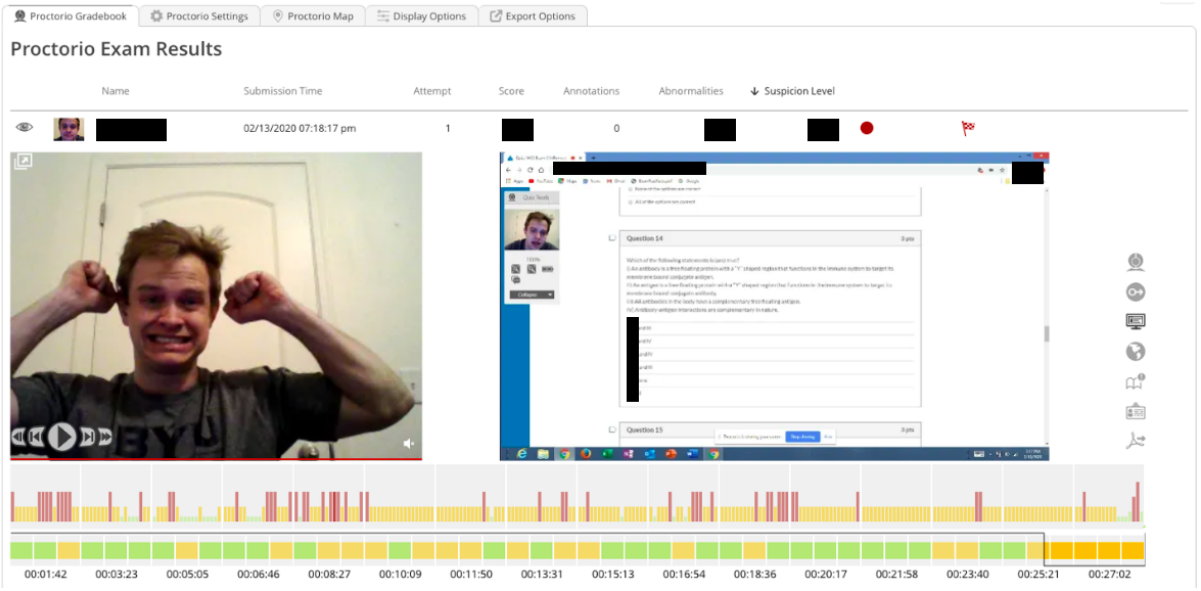

Fig. 8: A snapshot from the Proctorio AI proctoring results

An analysis of the "academic integrity" tickets chosen for a particular time period reveals the following:

- Academic integrity related tickets form about 5% of the total tickets: it is not the most pressing issue amongst students or at Outlier
- Within academic integrity, there are 3 major types of tickets: "notice", "warning", and "violation". Notices are typically for use of unauthorized alternatives (e.g. using Google calculator instead of the approved calculator). Warnings are typically for failure to upload an ID card. Violations are usually related to infractions such as searching online for an answer.
- Nearly 66% of the tickets are of the "warning" variety, "notices" form about 25% of the tickets, and "violations" make up the rest of the 9%

This shows that a major pain point within academic integrity for Outlier is easy and sound authentication, albeit academic integrity itself does not constitute an urgent need. Notices can be minimized by clearer communications, but it would be hard to eliminate them, as we are often

habituated to using certain tools, and in a stressful situation, might revert to a habit in spite of the rule. Violations or “cheating” is certainly not the biggest problem. It is also important to note that many a time, tickets tagged as “violation” lead to an investigation where the student is not found to have violated the rules. In order to further validate, I interviewed a student success team member who handles the academic integrity tickets, and they said that their ballpark estimate is that 1 in 200 students actually cheats. While there is no universal cheating rate to compare this to, Watson and Sotille (2010) found in a survey of 635 students that around 32% of the students admitted to cheating in both live and online examinations, and 5% of the live classroom, and 2% of the online classroom students said that they had been caught while cheating. This points to the need of creating a metric around academic integrity internally, and tracking it over time.

Based on interviews and analysis of student tickets, the following student needs were identified with respect to the examination experience:

- Minimize technical issues on the day of the examination, such as ID upload failure
- Reduce stress, uncertainty, and distractions during the examination
- Have multiple opportunities to showcase learning

These are the needs of the accreditation agency for distance education, under which category Outlier falls:

- Verify test-taker identity
- Protect student privacy

From the point of view of the student success team, which handles all the workflows related to academic integrity, the following are the needs with respect to the examination experience:

- Reduce time spent on detecting cheating
- Have standard operating procedures for detecting cheating
- Create an atmosphere of responsibility and honesty

The overarching Key Performance Indicator for Outlier that should bring together the above is that of student success.

Analysis of the Solution Space

To begin with, it is useful to see how AI proctoring delivers on the core needs at Outlier. Here is a table that summarizes how AI Proctoring performs against the above needs:

Need	**AI Proctoring Performance**
Minimize technical issues on the day of the examination	The onus of having all the required technology for successful proctoring is shifted to the student (smooth internet connection, updated browser version, webcam etc.). In addition to this, small issues such as eyeglass glare can trip the proctoring technology.
Reduce stress, uncertainty, and distractions during the examination	Studies show that AI proctoring has harmful psychological effects. As Kharbat et al (2021) surmised in their study, “Ultimately, the corresponding psychological concerns over

	being watched by a webcam contributes to students' feelings of fear and stress".
Have multiple opportunities to showcase learning	N/A
Verify test-taker identity	Certain AI proctoring softwares rely on facial recognition technology, which fails on dark skin especially in low light. Others, like Proctorio, do not perform this verification as they do not use FRT, and pass the task to the client.
Protect student privacy	Even if the AI proctoring software follows FERPA laws, it is unclear what happens to the student data that they send for review to the client. This passes the buck to the client. In addition to this, students feel that webcam monitoring in their homes is a violation of their privacy (Kharbat et al, 2021)
Reduce time spent on cheating detection	AI proctoring software uses imperfect technologies like gaze detection and FRT, and the brittleness is revealed in the fact that there

	are many false positives, with only 5% to 8% of the flags being actual violations in Outlier's case.
Have standard operating procedures for detecting cheating	AI proctoring algorithms are proprietary, and in providing the flags, the vendors provide no explanation as to why the flags were thrown up. Thus the task of setting rules around judging the flags falls on the client. This could potentially be a loophole as well if there are no SOPs, as human decision-makers are inherently biased.
Create an atmosphere of responsibility and honesty	As discussed above, these softwares create an atmosphere of stress and intrusion for the test-taker, and not a positive one. Besides, the link between stress/ anxiety and responsibility/ honesty is tenuous.

Several researchers have pointed to assessment design as a way to tackle the issue of academic integrity (Sullivan, 2016; Cluskey Jr. et al, 2011; McAllister and Watkins, 2012). Research has also found that establishing the institute's norms and frequently and prominently displayed honour code (Nguyen et al, 2020) also improves academic integrity overall.

With respect to online learning specifically, researchers have pointed out the need for thoughtful design and administration of formative and continuous assessment such that instructors and institutes are equipped with insightful data and analytics on the student's progress (Conrad and Openo, 2018). This could help to achieve 4 important things:

A. provide students with multiple opportunities to score points, i.e. build redundancy,
B. generate a pattern of performance that can be utilized for better student support,
C. help establish and improve on the correlation between proctored and unproctored test scores,
D. reduce the dependence on high-stakes testing to reveal a student's level of learning, and therefore,
E. make cheating impractical

In distance learning, authentication will continue to pose a challenge. At scale, this can be delivered using a combination of ID upload and facial recognition, password/ one-time password, and verification questions/ codes.

Thus, to summarize the above, the solution space consists of the following:

1. Making academic integrity visible and measurable
2. Administering the honour pledge creatively, frequently
3. Devising user verification/ authentication for high-stakes examinations
4. Increasing the number and type of avenues to showcase learning
    a. Variety of short quizzes: multiple choice questions with one or multiple correct options, fill-in-the-blanks questions, true/false, multiple drop-down etc.

b. E-portfolio

Conrad and Openo (2018) define e-portfolio as, "a collection of parts, often called 'artifacts,' that has been constructed or compiled by learners wishing to demonstrate their competence in a certain area... Technology has accelerated portfolio popularity and purpose by creating many different platforms accessible for users who have no particular design skills". The following is an example of suggested assessment criteria for an e-portfolio from University of Wisconsin Stout, from their distance learning recommendations[5]:

EPortfolio (Digital Portfolio) Rubric

Criteria	Unsatisfactory	Emerging	Proficient	Exemplary	Rating
Selection of Artifacts	The artifacts and work samples do not relate to the purpose of the eportfolio.	Some of the artifacts and work samples are related to the purpose of the eportfolio.	Most artifacts and work samples are related to the purpose of the eportfolio.	All artifacts and work samples are clearly and directly related to the purpose of the eportfolio. A wide variety of artifacts is included.	
Descriptive Text	No artifacts are accompanied by a caption that clearly explains the importance of the item including title, author, and date.	Some of the artifacts are accompanied by a caption that clearly explains the importance of the item including title, author, and date.	Most of the artifacts are accompanied by a caption that clearly explains the importance of the item work including title, author, and date.	All artifacts are accompanied by a caption that clearly explains the importance of the item including title, author, and date.	
Reflective Commentary	The reflections do not explain growth or include goals for continued learning.	A few of the reflections explain growth and include goals for continued learning.	Most of the reflections explain growth and include goals for continued learning.	All reflections clearly explain how the artifact demonstrates your growth, competencies, accomplishments, and include goals for continued learning (long and short term).	
	The reflections do not illustrate the ability to effectively critique work or provide suggestions for constructive practical alternatives.	A few reflections illustrate the ability to effectively critique work and provide suggestions for constructive practical alternatives.	Most of the reflections illustrate the ability to effectively critique work and provide suggestions for constructive practical alternatives.	All reflections illustrate the ability to effectively critique work and provide suggestions for constructive practical alternatives.	
Citations	No images, media or text created by others are cited with accurate, properly formatted citations.	Some of the images, media or text created by others are not cited with accurate, properly formatted citations.	Most images, media or text created by others are cited with accurate, properly formatted citations.	All images, media or text created by others are cited with accurate, properly formatted citations.	

Fig. 9: An example of a grading rubric for an e-portfolio

[5] Accessed from University of Wisconsin Stout website on may 8,2021 https://www.uwstout.edu/academics/online-distance-education/online-professional-development/educational-resources-rubrics/creating-and-using-rubrics-assessment

c. Cheat-sheets

Erbe (2007) found that requiring creation of crib or cheat sheets prior to an examination deepened student learning by providing structure to the learning, and also lowered text anxiety amongst the students.

d. Wikis: Student-maintained microsites relating to the course topics

e. Moderated group discussions: clear rubrics such as below from Swan et al (2013) can be used:

Discussion Rubric:
Each discussion post is graded according to the following rubric.

Points	Interpretation	
4	Excellent (A)	The comment is accurate, original, relevant, teaches us something new, and well written. Four point comments add substantial teaching presence to a course and stimulate additional thought about the issue under discussion
3	Above Average (B)	The comment lacks at least one of the above qualities, but is above average in quality. A three point comment makes a significant contribution to our understanding of the issue being discussed.
2	Average (C)	The comment lacks two or three of the required qualities. Comments which are based on personal opinion or personal experience often fall within this category.
1	Minimal (D)	The comment presents little or no new information. However, one point comments may provide important social presence and contribute to a collegial atmosphere.
0	Unacceptable (F)	The comment adds no value to the discussion.
No penalty	Excellent Subject	The subject field contains the main point of the comment. The reader clearly understands the main point of the comment before reading it.
1 point penalty	Minimal Subject	The subject field provides key word(s) only. The reader knows the general area that the comment deals with.
2 point penalty	Subject Field is Unacceptable	The subject field provides little or no information about the comment.

Fig.10: An example of a rubric for grading discussions from Swan et al (2013)

5. Diversifying the administration of examinations

a. Randomize questions and the order of the questions from multiple questions banks for each student

b. Frequency: Administering short quizzes often

c. Refresh question banks at a certain cadence with new questions, multiple versions of the same questions generated by changing numerals, wording

d. Flexible, timed examinations that don't allow an unfair advantage: allow students to answer anywhere between 20 and 40 questions in a fixed period of time, say 1.5 hours. Award full marks for a minimum number answered correctly, award extra points for answering more than minimum questions, allow carrying forward of leftover time on examinations to the next quiz. Timed quizzes have long been used as a way to further restrict cheating, as students would be forced to focus on completion and submission of correct answers. However, these are considered less optimal from the point of view of accessibility (7 out 10 accessibility accommodation requests at Outlier are regarding additional time on examinations). Thus, with the system of minimum requirement of quiz completion within a set time, and the opportunity to pick up extra points by answering extra questions looks to satisfy the criteria for timed examinations as well as accessibility.

6. Applying "backward design" to map assessments to learning objectives clearly so as to clarify expectations, minimize learner ambiguity and resultant frustration (complaints such as not having encountered a question or concept in the syllabus)
7. Providing the right analytics support to capture data and insights from the above
8. Creating an assessment and evaluation strategy that that can be benchmarked, measured, improved over time

Solution Implementation Strategy

Implementation of the above solutions will prove to be a challenge without a standard feature release project plan. In the next section, I provide the phase-wise implementation of the solutions.

Phase 0: Setting the KPI

In this phase, Outlier will set a measurable KPI against academic integrity, within student performance. Given that the crux of the solution lies in the connection between assessment design and academic integrity, the two have to be measured and improved together. The KPI will consist of the inputs/ outputs below, and an initial baseline will be determined and tracked thereafter:

1. Rate of academic dishonesty: Conduct an in-depth analysis of the data from the proctoring software and the ticket-resolution.
2. Student perception: user experience research on academic integrity and assessments, which draws from the Academic Dishonesty Scale and Net Promoter Analysis
3. Time spent: Aggregate and monitor time spent on proctored examinations
4. Student score: on the unproctored as well as proctored examinations
5. Time spent in cheating detection/ resolution

Phase 1: Redesign 1.0

In this phase, a redesigned honour pledge will be administered for one cohort in one course. The following is a suggested redesign:

a. At the start of the course, student reads the Academic Integrity Code

b. Student is required to type out and submit a simple honour pledge
c. At the time of joining the discussion board, student success team lays down the community norms, which includes academic integrity, and students are encouraged to respond with emojis confirming that they have read the same
d. At the 10-day mark, student is shown their honour pledge, and they respond with an emoji/ click "agree"
e. Few days prior to when typically a student might take the mid-term, student success coordinator re-ups their community norms message in the discussion board
f. Prior to taking the mid-term, the student is shown their honour pledge.
g. Thus, a student is constantly reminded of the academic integrity code and their own honour pledge

The second part of this redesign is in terms of the assessment. For a single chapter of a course, say the final chapter, assessment strategy will be built around the ideas highlighted in the solution space section. A combination of types of questions and administration will have to be picked, based on learner-centred pedagogy for that course. This will help to calibrate the instrumentation required for data capture.

It is unlikely that the KPI will show a major shift in this phase, as only a partial implementation of the solution has taken place.

Phase 2: Redesign 2.0

The assessment strategy from Phase 1 will be revisited, and student feedback will guide the process of scaling the assessment redesign to a whole course. It is important to note that at this point, we

still will continue to have AI proctoring, as ultimately, we want to show that the performance, time taken, student response on academic integrity surveys is the same across unproctored and proctored scenarios.

In this phase development of an in-house authentication is recommended, which could include the following:

1. ID upload by the student + verification at the start of the course
2. Single sign-on for examinations using personal/ institute credentials (which a student is unlikely to share with another person)
3. Photographic verification before starting the examination (photo taken via laptop / webcam)

It is important to continuously monitor the academic integrity KPI, and student feedback, in order to find blind-spots and avenues for improvement. This will contribute towards establishing the quality and integrity of the credits given by Outlier as well.

Example: Microeconomics

In order to help visualize the above, I take the example of Microeconomics. Beginning from Phase 1, students from one or two cohorts(depending on data needs: a sample size of at least 200 students) will be randomly assigned to a control and a treatment group. The control group will take the usual proctored midterm examination. The treatment group will take an unproctored midterm examination. Both the groups will be informed at the start of the course.

A learner signs up for, and signs into the Outlier LMS, after having purchased the Microeconomics course.

Current Workflow	**Pilot Workflow**
No ID upload on Outlier	Widget generated on student dashboard named "To Do List" which leads to site for ID upload (could be current Resources microsite)
Academic Integrity, Course details under Resources	Academic Integrity also on the same page

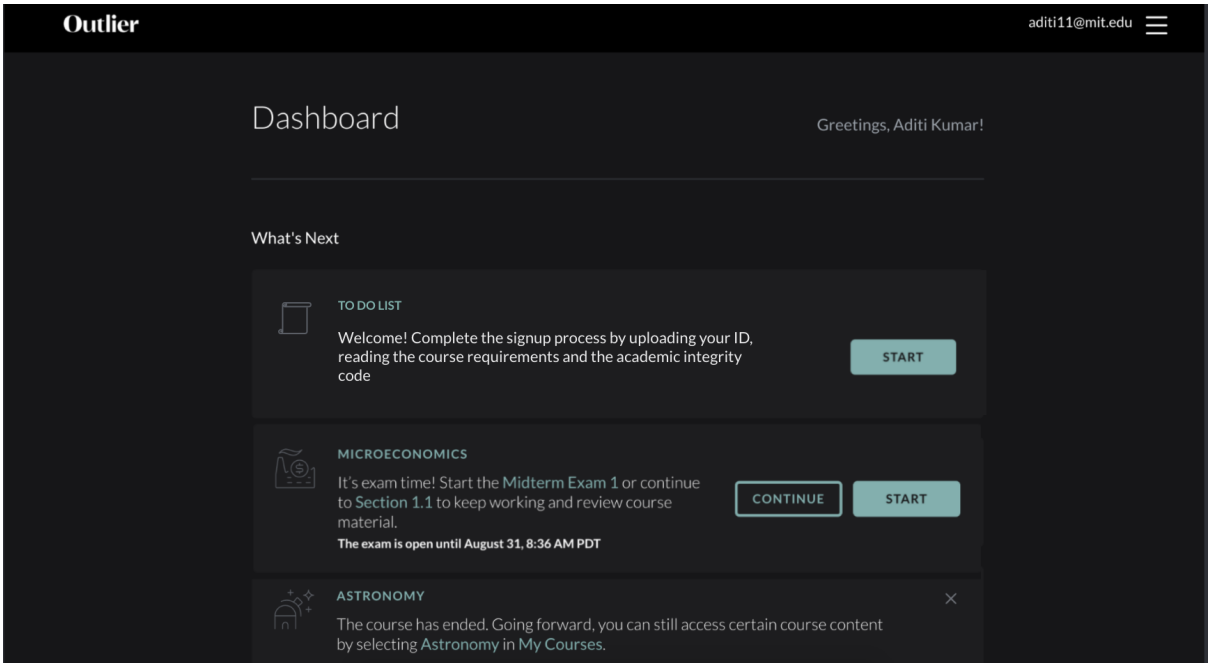

Fig. 11: Mock-up of the student dashboard with to-do list widget

Upon clicking on "Start" in the To Do List widget on the Student dashboard, the student is taken to what currently is the Resources microsite. Here, students can take up each activity one by one. A suggested order is :

1. ID Upload

2. Syllabus Overview (learning outcomes, topics etc.)
3. Course Administration Overview (duration, timelines, assessment structure etc.)
4. Academic Integrity Code (existing Outlier code of conduct)
5. Personalized Honour Pledge (similar to signing a contract, something along the lines of "I, <name>, hereby confirm that I have read the Academic Integrity Code, and pledge to not resort to unfair means and unauthorized resources or help to complete any part of the coursework" and the student types in their name/ date on this)

Upon completion, the student can navigate back to their Dashboard.

The student opens the Microeconomics course site.

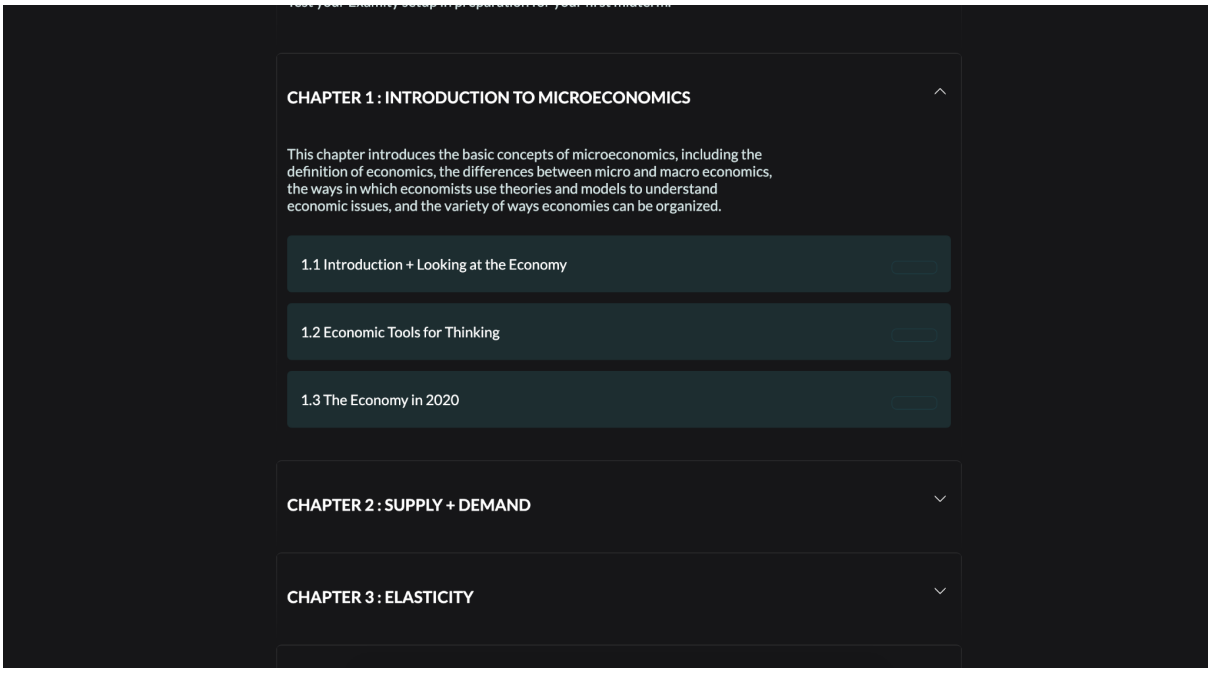

Fig. 12: The course site for Microeconomics

From here, they navigate to the first sub-chapter.

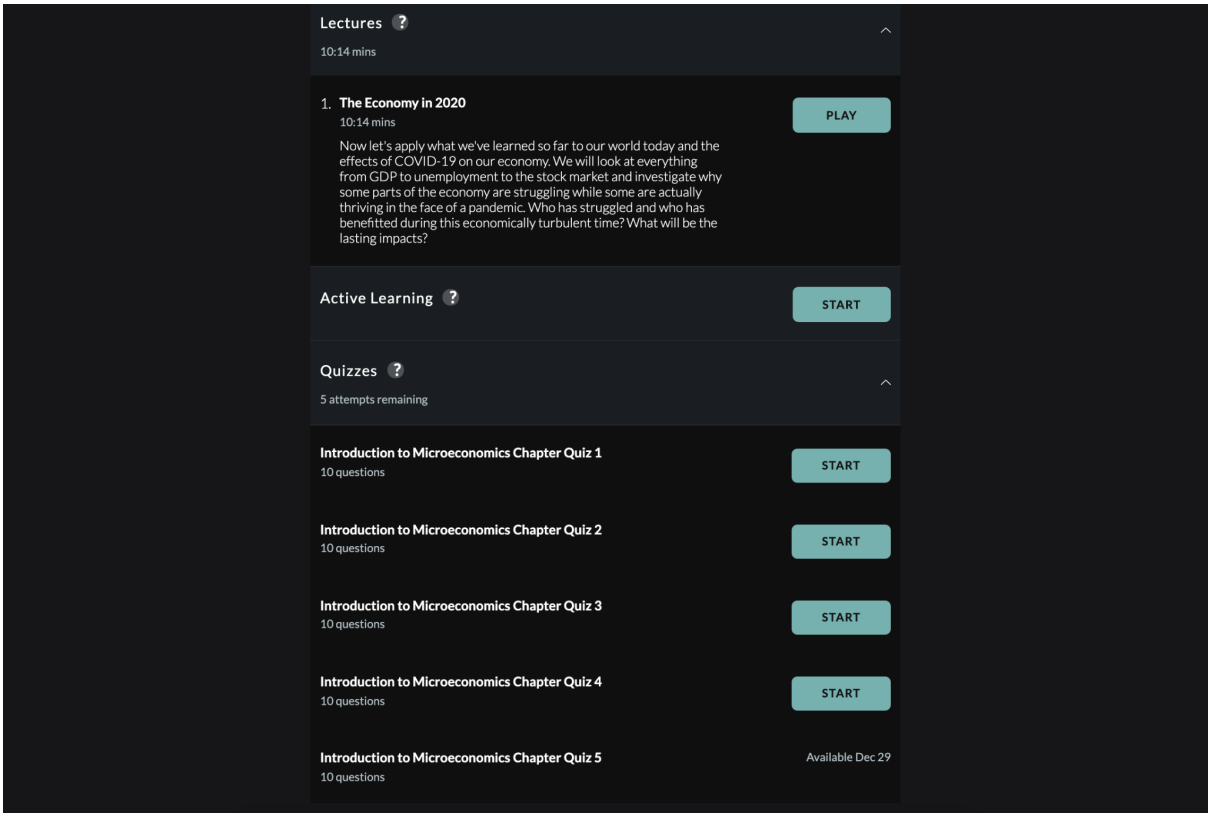

Fig. 13: All the quizzes at the end of the final sub-chapter of Microeconomics

Current Workflow	**Pilot Workflow**
Each sub-chapter consists of videos, e-textbook with a few learning checkpoints, and ungraded practice problems. All the quizzes are at the end of the last sub-chapter, within the first chapter.	One short quiz (~5 to 8 minutes) in each sub-chapter. Varying formats like fill-in-the-blanks, multiple choice with 1 or more correct options.

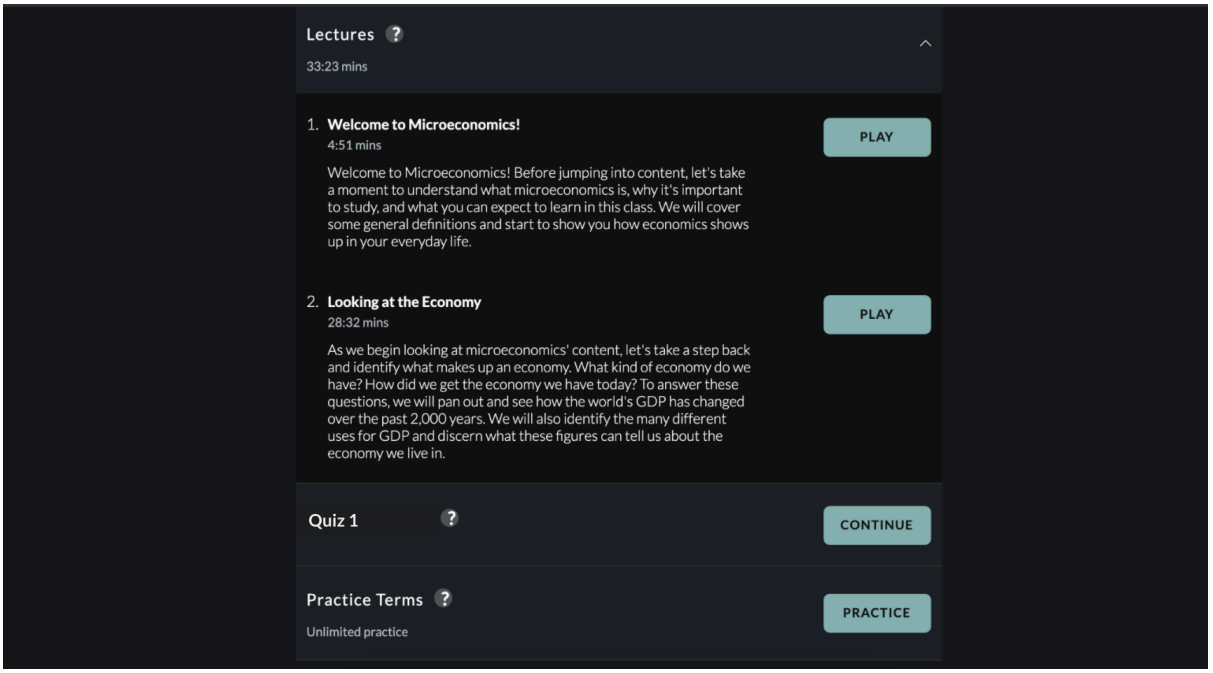

Fig. 14: Mockup of Quiz included in the first sub-chapter

After the completion of all sub-chapters of the chapter, a final quiz fashioned like the flexible, timed quiz mentioned in the solution space discussion earlier in the dissertation will be implemented.

Given that there are 5 chapters prior to the first Microeconomics midterm, the assessment strategy will have to be thoughtfully built out for the 5 chapters, in an integrated fashion, i.e. ideally, the different parts of the assessment must build on one another, and not be disjoint. This can be done by applying "backward design" first for the 5 chapters taken together, and then broken down further.

In the pilot study, creation of cheat-sheets as an activity can also be inserted for the midterm. Students will have to be informed upfront. While this is out beyond the scope of the pilot study perhaps, one way to do this is by allowing the students to annotate their learning, as a supplemental

process towards creation of a cheat-sheet. Pasted below is an example of how Coursera currently offers this functionality, by allowing students to highlight the video transcript[6]:

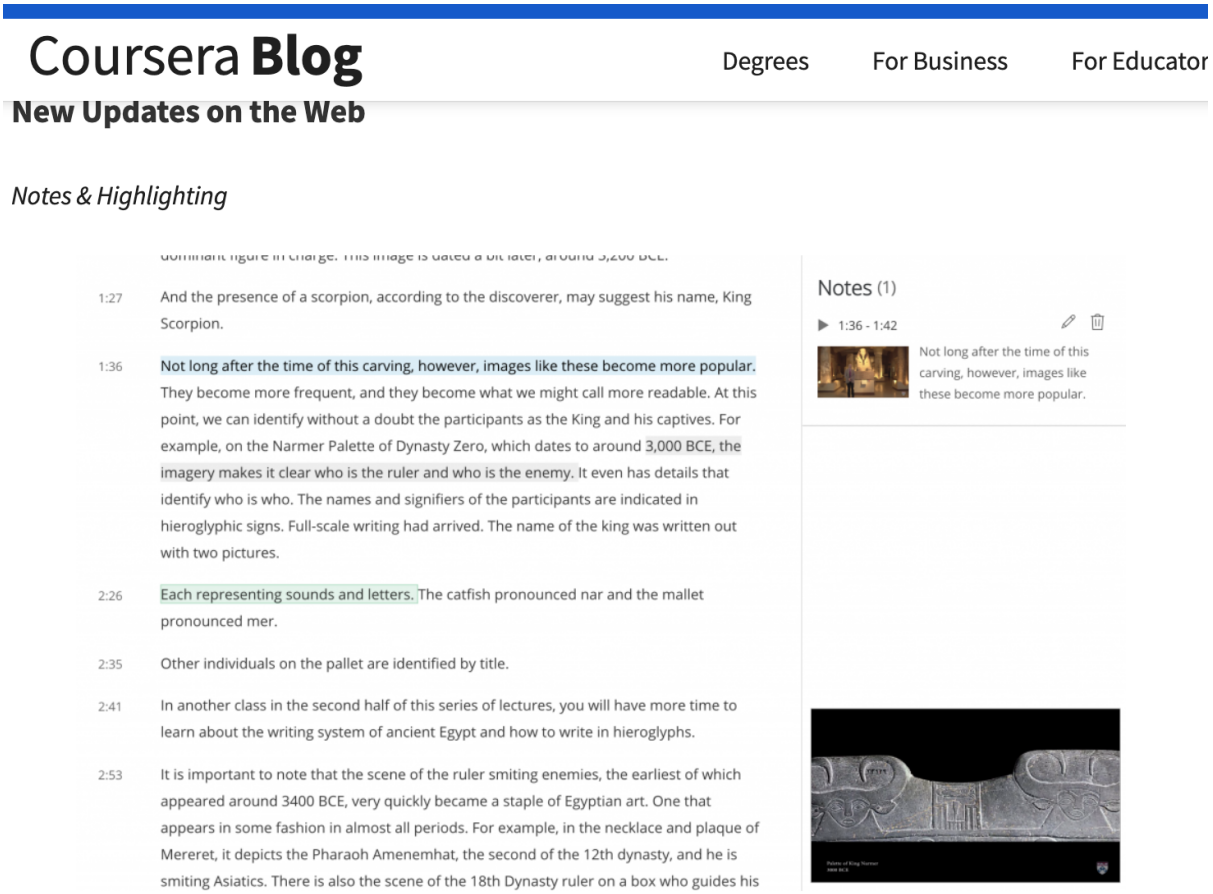

Fig. 15: image of Coursera blogpost about highlighting within transcript

Students can be asked to click a picture of their cheat-sheets and upload them for an extra point.

In addition to all of the above, the student discussion board should be leveraged by the student success team to talk about academic integrity. As indicated in the dissertation in section 6.2, the student's personal honour pledge should be surfaced before the midterm, or at a frequency deemed appropriate after discussion and testing.

[6] Accessed from Coursera Blog on May 12, 2021 (https://blog.coursera.org/updates-to-your-learning-experience-on-coursera/)

As discussed in the beginning of this section, the midterm itself will be proctored for some students and unproctored for some others, based on the random assignment to control and treatment. At all times, the academic integrity KPI metrics need to be monitored for each student for the study. At the end of the study, the metrics will be analysed for and compared across the control and treatment groups.

Conclusion

In this design assignment, I set out to recommend design alternatives to AI proctoring software for online examinations for an education technology startup, Outlier.org. I first conducted interviews and analysed company data to determine the needs associated with examination experience design. I then considered how AI proctoring software maps to these needs, and found that it did not fulfill, and in some cases went in the opposite direction of, the needs identified. Through interviews and secondary research, I identified the solution space, which included assessment design strategy, authentication, and academic integrity promotion. I then laid out a phased approach to implementing the solutions, including defining the key performance metrics and ways to measure the same.

AI proctoring softwares often rely on brittle technologies that produce a number of false flags, biased against dark skin, poor lighting, unstable internet, and therefore are in direct opposition with the principle of equity. Moreover, the burden of shortcomings of the poor technology is passed to clients (too many flags to review) and test-takers (software doesn't work well in low light

or poor connectivity). In addition to this, relying on third party software is unsustainable in the long run for a startup.

However, it is to be noted that part of the solution, namely assessment design strategy, can never be final or foolproof, and as user behaviour and needs evolve, the system must keep up. Therefore it is vitally important to make the problem measurable, and tie it back to the business and organization goals.

References

Brigitte Erbe (2007) Reducing Test Anxiety While Increasing Learning: The Cheat Sheet, *College Teaching, 55:3*, 96-98, DOI: 10.3200/CTCH.55.3.96-98

Cluskey, Jr., G.R., Ehlen, C.R., & Raiborn M.H. (2011). Thwarting online exam cheating without proctor supervision. *Journal of Academic and Business Ethics, 4*, 1-7

Conrad, D., & Openo, J. (2018). Assessment strategies for online learning: *Engagement and authenticity*. Athabasca University Press.

Erwin, T. D. (1991). Assessing Student Learning and Development: A Guide to the Principles, Goals, and Methods of Determining College Outcomes.

Holden, O., Kuhlmeier, V. A., & Norris, M. (2020). Academic integrity in online testing: A research review.

Hussein, M. J., Yusuf, J., Deb, A. S., Fong, L., & Naidu, S. (2020). An evaluation of online proctoring tools. *Open Praxis, 12(4)*, 509-525.

Kharbat, F. F., & Daabes, A. S. A. (2021). E-proctored exams during the COVID-19 pandemic: A close understanding. *Education and Information Technologies,* 1-17.

McAllister, C., & Watkins, P. (2012). Increasing academic integrity in online classes by fostering the development of self-regulated learning skills. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas, 85(3)*, 96-101.

Nguyen, J. G., Keuseman, K. J., & Humston, J. J. (2020). Minimize Online Cheating for Online Assessments During COVID-19 Pandemic. *Journal of Chemical Education, 97(9)*, 3429-3435.

Selinger, E., & Leong, B. (2021). The Ethics of Facial Recognition Technology. Forthcoming in *The Oxford Handbook of Digital Ethics* ed. Carissa Véliz.

Sullivan, D. P. (2016). An Integrated Approach to Preempt Cheating on Asynchronous, Objective, Online Assessments in Graduate Business Classes. *Online Learning, 20(3)*, 195-209.

Swan, K., Shen, J., & Hiltz, S. R. (2006). Assessment and collaboration in online learning. *Journal of Asynchronous Learning Networks, 10(1)*, 45-62.

Watson, G. R., & Sottile, J. (2010). Cheating in the digital age: Do students cheat more in online courses?

Wiggins, G. P., Wiggins, G., & McTighe, J. (2005). *Understanding by design*. Ascd.

Appendix A

The interview questions came from the following pool of questions, and differed based on the audience. The interviews were conducted on Zoom, and recorded. The transcripts were then analysed.

1. When did Outlier first consider proctoring? What was the context?
2. How, why was <proctoring software platform name> selected?
3. Was there a solution, or a set of solutions, that Outlier used prior to/ in conjunction with proctoring?
4. What is currently the experience of working with the current AI proctoring platform? What works, what doesn't? Two to three points should suffice
5. What is the current workflow for proctoring at Outlier?
    a. What happens after an exam?
    b. Who all has access to the recordings? For how long?
    c. What kind of data & insights does the AI proctoring platform share?
    d. With whom does the AI proctoring platform currently share?
    e. What happens if the AI proctoring platform flags an event?
    f. How is cheating determined?
    g. What happens after cheating is confirmed?
    h. Does Outlier share the information of the confirmations with the AI proctoring platform?
    i. Some examples or instances of the above workflow
6. Is there a dedicated team for this workflow? Who are the people involved?
7. Have the cheating instances been documented?
8. What has the feedback from a) students, b) student support, c) internal team, d) any other relevant stakeholder been about the AI proctoring platform experience?
9. When did you start thinking about assessment design and cheating deterrence?
10. What have your observations been with respect to students in an online learning setting?

11. What are the best practices you follow for online learning?
12. Who are the key stakeholders to keep in mind while considering academic integrity system requirements?
13. What are the top 3 to 5 feedback trends from students?
14. What are some of your reflections on academic integrity at Outlier?
15. Can you share more about the academic integrity issues at Outlier and the break-up of the same?
16. Can you share more about the instances of serious academic integrity violations?
17. Have there been complaints about proctoring itself from students?
18. According to you, what are the top challenges with respect to academic integrity?
19. Have you given thought to how you might tackle issues around exam experience/ academic integrity, and if yes, would you like to share?
20. What is the process for design of assessments? Do you do agile sprints, consultative workshops?
21. Have you implemented any major changes to assessment design in the past? If yes, how was this done?
22. How do you handle student tickets, especially those related to technical difficulties during exams?
23. How much of student progress and activity is currently tracked on the website for analytics?
24. What according to you are the foundational principles for assessment that you follow? Can you please also share your process?
25. Have you revisited the assessment structure or design in moving from in-person to online instruction?
26. Do you have a strategy to refresh/ change questions banks?
27. What have been some of your key takeaways in teaching online?
28. What are some of the steps you have taken to ensure academic integrity?
29. Do you rely on the LMS for any assessment administration/ grading/ creation? What are the top 2 things that need improvement, and top 2 things that are done really well on the LMS that you currently use?

Appendix B

AI Proctoring Platform Product Offering Questions

The following are Yes or No questions regarding the AI proctoring platform, and the features in use:

(modified from https://files.eric.ed.gov/fulltext/EJ1285031.pdf)

Proctoring Features

- Live human proctors available
- Internet required
- Secure/encrypted transferring of data
- Student able to book exam time
- Training provided
- Proctoring provider certified
- Students can interact with proctors
- Student can message issues to proctors
- Students get live exam instructions
- Proctor able to see students screen
- Stop proctor to view students screen
- Recorded video reviewing option
- Pause test/ cancel test
- Automated proctoring
- Keystroke checking
- Audio recording
- Browser lockdown
- Authentication option
- Web camera needed
- Log reports
- recording storage option
- Test review option

- Incident logs with date & time
- Customising options for institution

Lockdown Features

- Available on both Windows and Mac
- Plugin for browser
- Avoids control options on the browser
- Stops navigation (forward/back)
- Stops concurrent tests
- Stops right clicks using mouse
- Stops printing
- Hides taskbar
- Hides desktop
- Stops minimising window
- Stops maximising window
- Stops copying & pasting
- Stops other applications
- Stops starting of other applications

Authentication options

- User required to authenticate
- Username provided/required
- Password provided/required
- Student ID required
- Keystroke analytics
- Ability to do facial recognition
- Ability to do voice recognition
- Fingerprint scanning required
- Iris scanner required/available

Webcam Features

- Web camera required
- Room panning allowed